

Book bans in political context: Evidence from US schools

Marcelo S.O. Goncalves^{a,1}, Isabelle Langrock^{b,c,1}, Jack LaViolette^{d,e,1} and Katie Spoon^{f,*1}

^aDuke University, Sanford School of Public Policy, Durham, NC 27708, USA

^bAnnenberg School for Communication, University of Pennsylvania, Philadelphia, PA 19104, USA

^cSciences Po, Centre for Research on Social Inequalities, Paris 75007, France

^dDepartment of Sociology, Columbia University, New York, NY 10027, USA

^eIncite, Columbia University, New York, NY 10027, USA

^fDepartment of Computer Science, University of Colorado Boulder, Boulder, CO 80309, USA

*To whom correspondence should be addressed: Email: katherine.spoon@colorado.edu

¹All authors contributed equally to this work.

Edited By: Morris Levy

Abstract

In the 2021–2022 school year, more books were banned in US school districts than in any previous year. Book banning and other forms of information censorship have serious implications for democratic processes, and censorship has become a central theme of partisan political rhetoric in the United States. However, there is little empirical work on the exact content, predictors of, and repercussions of this rise in book bans. Using a comprehensive dataset of 2,532 bans that occurred during the 2021–2022 school year from PEN America, combined with county-level administrative data, multiple book-level digital trace datasets, restricted-use book sales data, and a new crowd-sourced dataset of author demographic information, we find that (i) banned books are disproportionately written by people of color and feature characters of color, both fictional and historical, in children's books; (ii) right-leaning counties that have become less conservative over time are more likely to ban books than neighboring counties; and (iii) national and state levels of interest in books are largely unaffected after they are banned. Together, these results suggest that rather than serving primarily as a censorship tactic, book banning in this recent US context, targeted at low-interest children's books featuring diverse characters, is more similar to symbolic political action to galvanize shrinking voting blocs.

Keywords: book bans, politics, education, censorship, polarization

Significance Statement

Book banning is increasingly common in US schools. While most studies focus on centralized, state-sponsored censorship, individuals such as parents and local organizations have participated in this recent wave of banning. Our study empirically describes banned books and authors, finding high rates of children's books written by authors of color among banned books. Furthermore, we analyze the local contexts that predict bans and evaluate how interest changes after books are banned. In sum, we suggest that this wave of book bans is best understood as a form of political action in increasingly contested local contexts rather than as a means of effective censorship. These findings contribute to scholarship on social movements, polarization, and censorship in contemporary democracies.

Introduction

While a quintessential signifier of censorship and intellectual suppression, book banning is not a foreign practice to the American public (1). United States schools and libraries have banned books with some regularity for the past two centuries, as traditional norms were challenged by modernist and scientific thought (2, 3). However, the 2021–2022 school year saw a drastic increase in book bans across the country, often through mandates from school boards and parent complaints (4, 5). Following the 2020 murder of George Floyd and the intensification of a partisan

“culture war” (6, 7), book bans have become central to a broader conversation around politics, civics, and identity.

Journalists have diligently documented the recent rise in book bans, particularly noting how bans directed against profanity, violence, and sexual content target books with LGBTQ+ and Black characters (8–11). While there are cases, most notably around the work of Mark Twain, where books are removed from the curriculum or annotated to note the historical context, the vast majority of bans follow larger debates about the inclusion of critical race theory (12), LGBTQ+ perspectives, and inclusive gender

Competing Interest: The authors declare no competing interest.

Received: November 10, 2023. **Accepted:** May 7, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of National Academy of Sciences. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

theory (13, 14) in school curriculums. To proponents of bans, exposure to books that convey these theories is a form of indoctrinating students, such that bans protect children from inappropriate content (15). By contrast, opponents describe bans as questionably legal attempts to deny young people access to information about the reality of systematic race- and gender-based discrimination in US public institutions and to vital social representations affirming a wide range of experiences and identities (16, 17). Bans seemingly censor particular identities exactly at the time that students begin to explore their own.

Academic research on contemporary book banning in the United States is scant but growing. Legal scholars have identified the contradictions between students' First Amendment rights and censorship attempts (17, 18), while library science scholars have described recent book bans as a revival of McCarthyism, diminishing intellectual freedom and a sign of increasing precarity for public libraries and schools (4, 19). Education scholars find little evidence that bans protect children (20) and further argue that bans, in infringing upon children's human rights and their ability to access information, are likely to hinder the development of critical thinking skills (16).

Outside of book bans, much of our understanding of contemporary information censorship comes from the study of authoritarian actors and online environments, where states take a variety of measures to suppress oppositional information (21–24). Yet unlike state-sponsored forms of information suppression, book bans in the United States exist within a framework of participatory democracy. Bans are supported by complex and often opaque collaborations between local parent organizations and national political organizations such as Moms for Liberty, with close ties to the Republican Party (25) and are adjudicated through the democratic operations of school boards. As book bans are dispersed across the country, what are the motivating factors uniting them? To what extent are they predictable, both politically and in regards to the content they target?

Our study answers these questions through a systematic analysis of 2,532 book bans that occurred in the United States during the 2021–2022 school year (26) that we annotate and substantially extend with administrative data, multiple digital trace datasets, restricted-use book sales data, and a new crowd-sourced dataset. These multifaceted data allow us to empirically assess the full spectrum of content being banned—the majority of which, we show, is written by women and people of color and features characters of color, both fictional and historical—but that otherwise does not neatly align with the descriptions of gratuitous sexual content or dogmatic texts on race and gender theory. We also assess the heterogeneity of socio-political contexts in which book bans occur, a level of detail crucial to understanding book bans as a form of collective action embedded within multiple layers of social context. Altogether, our findings suggest that it is perhaps more apt to think of current book bans as a political tactic to galvanize conservative voters in increasingly divisive electoral political districts, rather than as a pragmatic effort to restrict access to certain materials.

Additionally, we test for the presence of two competing common effects of censorship: (i) the successful suppression of information (27) or (ii) a backlash effect, also known as the “Streisand effect,” where attempts at censorship drive more attention to the information (28). We find that there is very little interest in banned books even before they are banned. Furthermore, we find that the bans rarely intervene to draw more or less attention to a book, with national and state levels of interest in books remaining largely unaffected after they are banned. These findings

suggest that while many banned books and authors are in line with the “culture wars” surrounding race and gender, bans are likely ineffective as a form of mass censorship of these topics. These findings compel us to reconsider book bans not solely as cultural or educational issues but as forms of political action, targeting the ballot box via library shelves and classrooms.

Results

We investigate three aspects of contemporary book bans. First, we assess the *variety of content and identities of authors* that are being targeted. We address the first question by grouping crowd-sourced book genres into broad thematic clusters, and the latter by collecting and analyzing self-identified author demographic information. Second, we ask *in what contexts* books are most likely to be banned via a series of regression analyses applied to a broad range of county-level demographic factors. Finally, we ask *how interest changes* after books are banned through a pre-post analysis across several indicators of interest, including book sales and Google searches. Table S1 lists each research question and data source.

Types of banned books and authors

We use an inductive, data-driven approach to produce a high-level typology of books that are banned. The goal of this approach is to identify high-level book groupings (which we refer to as “genres”) based on book subgenres, such that each book is more similar in subgenre composition to the other books within its genre than to books in other genres. We use crowd-sourced book subgenres from Goodreads—a popular website where users can list and review books they read—yielding 143 unique subgenres (e.g. “Fantasy,” “LGBT,” “American History”) among the banned books in our sample, with each individual book associated with up to 10 subgenres. This procedure, based on the commonly combined Uniform Manifold Approximation and Projection (UMAP) and Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) algorithms ((29, 30); see Materials and methods for more details on the clustering procedure), identified six thematic genre clusters that parsimoniously characterize the banned books (Fig. S1).

We summarize these genres, in order of frequency, as: (i) children's books with diverse characters, including both LGBTQ+ characters and characters of color (37%), (ii) nonfiction books about social movements and historical figures (22%), (iii) fantasy and science fiction (10%), (iv) young adult queer romance novels (10%), (v) women-centered fiction (10%), and (vi) fiction with mature, nonromance themes, like violence or drug use (7%), with 4% of books remaining unclustered as outliers (Fig. 1A). These trends remain relatively stable across the 12 months of the 2021–2022 school year, with a peak for book banning in the winter months, when school boards are more likely to be actively meeting (Fig. 1B).

In addition to characterizing the main genres targeted by book bans, we identify how bans also operate to censor authors from various demographic groups. Through an Amazon Mechanical Turk crowd-sourcing task, we collected the self-identified gender, race, ethnicity, and sexuality of the 1,139 unique authors in our sample (with 95.7% of authors having a publicly-available online profile containing the information). We found that 64% of banned authors are women and 3% are nonbinary, while only 29% are men. In addition, 19% of authors in our sample self-identify as LGBTQ+ and 39% as people of color (Asian, Black, Hispanic, Indigenous, or otherwise self-identifying as a person of color).

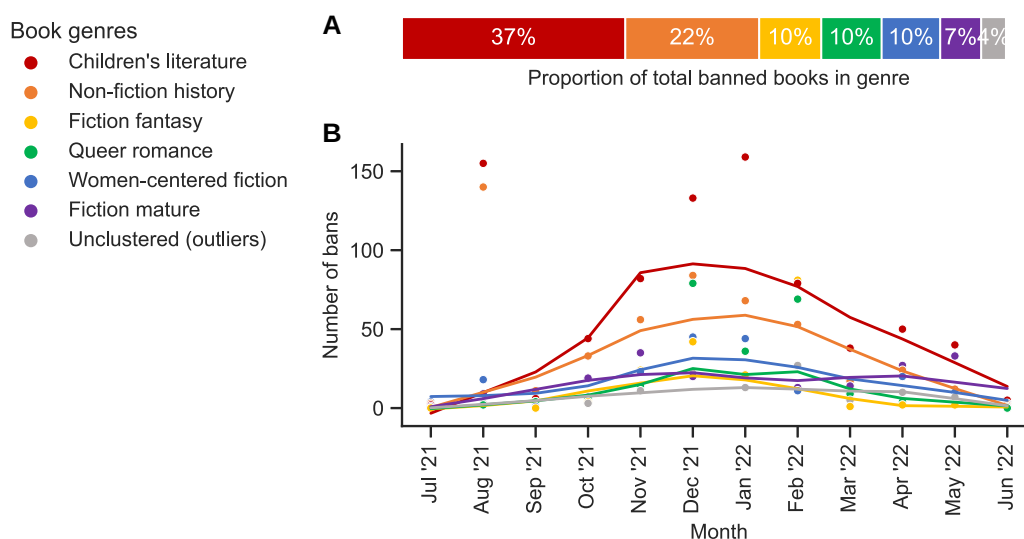


Fig. 1. Children's books featuring diverse characters are most likely to be banned. A) Proportion of banned books clustered into each genre. Books ($N = 1,370$) can only be clustered into one genre, so genres sum to 100%. B) Number of bans per genre over time. Number of bans ($N = 2,532$; books can be banned multiple times) per genre each month over the 2021–2022 school year, smoothed with loess.

To identify how the demographics of banned authors might systematically differ from the general author population, we compare them to (i) the US Census group who self-identify as authors (31) and (ii) a dataset of authors who wrote the most popular books from 1950–2018 (32), where popular books were defined as those published by the most prolific publishing houses and held by at least 10 libraries. We find that while women and LGBTQ+ authors are slightly overrepresented among banned authors, authors of color are strongly overrepresented among banned authors (women, $\chi^2(2, N = 4,631) = 14.5, P < 0.001$; LGBTQ+, $\chi^2(1, N = 4,610) = 6.8, P = 0.009$; people of color, $\chi^2(2, N = 4,887) = 839.6, P < 0.001$; Fig. S2).

In fact, the odds that an author of color was banned is 4.5 times higher than a white author, in comparison to all authors ($z = 7.8, P < 0.001$; Fig. 2A), and 12.0 times higher than a white author, in comparison to only the most popular authors ($z = 25.1, P < 0.001$; Fig. 2B). This phenomenon is driven largely by women of color, who make up 24% of banned authors in our sample, roughly twice the proportion of authors of color overall (31) and five times the proportion of authors of color who wrote the most popular books from 1950–2018 (32). Unfortunately, neither reference group of authors collected intersectional gender and race information (e.g. the proportion of authors who are women of color) or information about nonbinary authors (Fig. S2).

Further, we find that the types of books authors write are associated with their identities. Children's books and nonfiction books about social movements were the most popular genres for each intersectional group of authors (e.g. non-LGBTQ+ white men or LGBTQ+ women of color), with the exception of nonbinary authors, who were more likely to write fantasy sci-fi books than any other genre (Fig. S3). However, non-LGBTQ+ women of color were more likely than any other group to write children's books featuring diverse characters, the most frequently banned category of books. By banning children's books, women authors of color are effectively banned as well.

Socio-political environments of book bans

While the majority of book bans occurred in Florida, Pennsylvania, Tennessee, and Texas, they appear across the

country (32 states), indicating that there are contextual factors motivating book bans beyond simple regional tendencies (Fig. 3A). In order to assess the factors that predict book bans, we establish a comparison group comprised of counties that were not the site of book bans but which share a commuting zone with at least one county whose schools did ban books. Commuting zones are official designations developed by the US Department of Agriculture Economic Research Service that group counties based on dense economic interrelationships. Each county belongs to exactly one commuting zone, of which there are 709 in total. This empirical strategy allows us to investigate how counties with school districts that ban books might differ along social and political lines despite sharing a similar socio-economic background, regional context, and, presumably, culture.

Given the lack of prior quantitative research about the current wave of book bans, we test for a broad range of potentially associated factors including immigration patterns, average income and education levels, rates of religious observance, racial demographics, and political participation (Fig. S4). For example, one could imagine that racial threat (33, 34) associated with local influxes of nonwhite immigrants might increase the likelihood of local book bans, or that parents of higher socioeconomic status have more free time to devote to volunteer activities (35), or that religiosity net of political identity is associated with support for censorship (36).

Across all factors, one of the most substantial distinctions between counties that banned books and those in the same commuting zone that did not was the change in vote share won by Republican candidates in US presidential elections. From 2000 to 2016, there was no significant difference in Republican vote share between counties that banned books and others in their commuting zone that did not ban books (Fig. 3B). However, in 2020, counties with a weakened Republican majority, although still remaining above 50%, went on to ban books during the 2021–2022 school year while the nearby counties where the Republican majority gained strength did not ban books. Regression analyses identify that books are banned in communities that are wealthier, more educated and whiter, but the change in Republican vote share

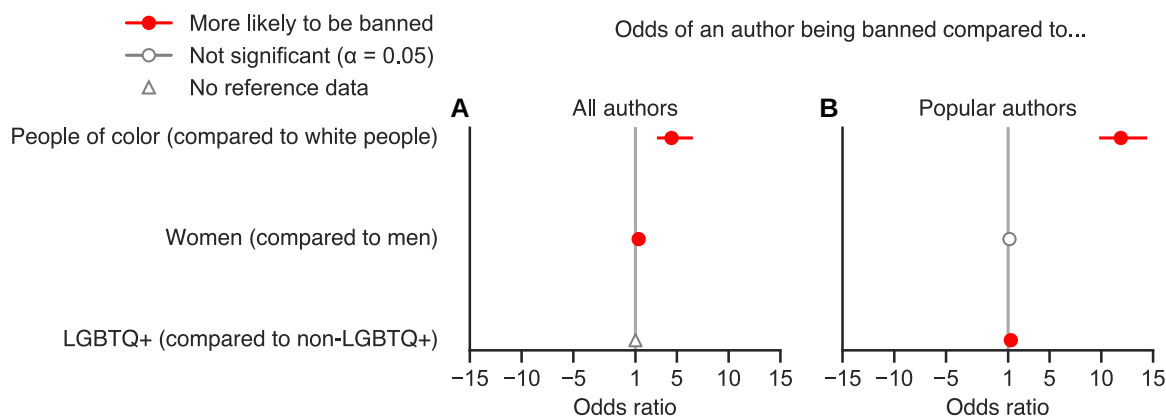


Fig. 2. Books written by authors of color are far more likely to be banned. Odds ratios, split by demographic variable (race, gender, and whether an author identified as LGBTQ+), comparing the proportion of authors who wrote banned books in the United States during the 2021–2022 school year to A) the proportion of authors who listed their occupation as a writer or author in the United States in 2022 (31), which does not collect data on LGBTQ+ authors, e.g. where $\text{odds}_{\text{POC}}/\text{odds}_{\text{white}} = (n_{\text{POC banned}}/n_{\text{POC all}})/(n_{\text{white banned}}/n_{\text{white all}})$ and B) the proportion of authors who wrote the most popular books in the United States from 1950–2018 (32), e.g. where $\text{odds}_{\text{POC}}/\text{odds}_{\text{white}} = (n_{\text{POC banned}}/n_{\text{POC popular}})/(n_{\text{white banned}}/n_{\text{white popular}})$. 95% confidence intervals and statistical significance were assessed via a z test.

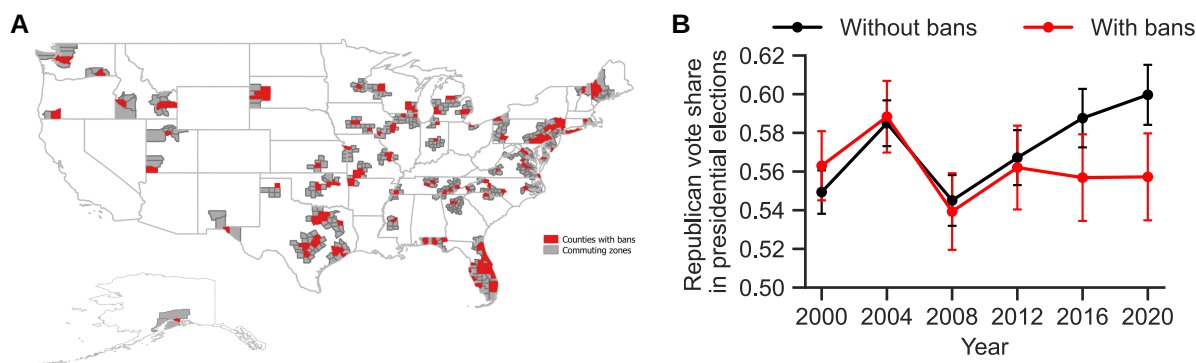


Fig. 3. Republican vote share predicts bans across counties in the same commuting zone. A) Distribution of book bans across US counties (2021–2022). Counties containing school districts that banned books in the 2021–2022 school year are compared to neighboring counties in the same commuting zones that did not ban books. There were no book bans in Hawaii (not displayed) during the study period. B) Republican vote share in book-banning counties vs. neighboring counties. The fraction of voters in counties with and without book bans who voted for the Republican presidential candidate each year, beginning in 2000.

remains one of the strongest and most significant predictors across multiple specifications (Table S2). In other words, Republican strongholds were not likely to ban books while counties with increasingly precarious conservative majorities were.

Interest in banned books

We use two different indicators—internet searches and book sales—to assess national interest in the banned books in the months prior to and proceeding each ban. Both interest indicators only cover a fraction of the total number of bans (Bookshop.org, 13%; Google Trends, 62%), with data unavailable for the remaining bans because of low interest (i.e. there were no sales or too few Google searches to populate the Google Trends data). The different rates of available data across the two indicators reflect the different types of interest captured in the two datasets; for example, it takes significantly less effort to search for a book online than it does to purchase it.

There is strikingly low overall national interest across both indicators throughout the period of our study. This is particularly noteworthy given the historical focus of censorship on banning

popular books. The low data availability is consistent with our other data collection efforts (see Section S3 for more information). The individuals and organizations that advocate for book bans presumably strive for a decrease in interest, which would be a sign of effective censorship, an effect we are unlikely to see at the national level. Conversely, we could expect increased interest due to a “Streisand effect” (28), whereby interest rises following the ban due to the increased media attention or as a form of protest.

We observe a small positive change—approximately 1%—comparing the three months following a ban and the 3 months prior among national Google search results for books (Fig. 4A), but this is not evident in the sales data (Fig. 4B). However, this is tempered by the large rate of missing data and the null results of the sales, suggesting that book bans produce little change in the number of people who engage (or do not) with a book. Primarily, we find that bans are directed at books that largely do not attract the public’s interest to begin with.

The relationship between national levels of interests in banned books and the local effects are unknown. Indeed, data availability prohibits more targeted estimates (Section S3). While national levels of Google searches increase slightly after books are banned, at

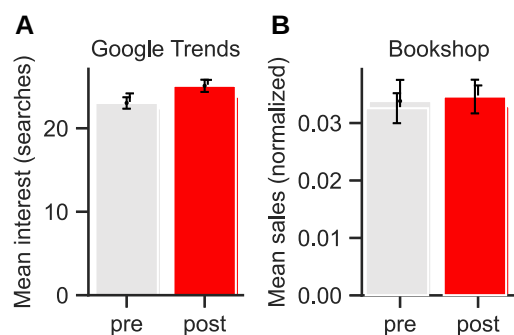


Fig. 4. Interest in books does not substantially change after they are banned. Average interest across the 3 months prior to each ban and 3 months after each ban, with 95% confidence intervals, for A) Google Trends searches, which has a small significant positive change in mean interest and B) [Bookshop.org](#) sales, which do not significantly change.

the state level, searches do not change significantly (Fig. S9). At the local level, Seattle Public Library's open data portal allows us to access book check-out data, and we find these local results to be consistent with the national- and state-level trends: interest is generally low both pre- and post-ban, and does not change (Fig. S10). However, no school district in Seattle banned books during our study period nor is the city representative of areas that generally ban books. Even so, we interpret these null results as confirming our broader argument that contemporary book bans do not generally target popular books.

Disaggregating interest data for each of the five most frequently banned books in our sample, we find that there is only a small increase in interest for one book: *Gender Queer: A Memoir* (37), which received more Google searches in the months after a ban than it did preceding (Fig. S5). It is not possible to distinguish the increased interest in *Gender Queer: A Memoir* as a backlash effect to the ban or a general rise due to the increased media attention the book received as the country's most frequently banned book.

Discussion

Book bans are increasingly common in US schools and libraries, suggesting censorship is growing within certain participatory democracy systems. Our large-scale study identifies consistent features of contemporary book bans: the books targeted for bans are most often children's books and nonfiction books about historical figures; they are disproportionately likely to be written by women and authors of color, particularly women of color; and the general US public has low levels of interest in them, both before and after bans occur. Further, we find books are more likely to be banned by school districts in counties with increasingly contested presidential elections compared with neighboring counties: specifically, those in which the Republican candidate, while still winning over 50% of votes, faces stronger competition from Democratic challengers than in previous elections. This is one of the strongest predictors that a school district within a county will ban a book. Despite the increasing prominence of book bans in American political and social life, bans tend to target books with relatively low sales and interest to begin with, suggesting that the goals of traditional forms of censorship (i.e. suppression of oppositional information) are not the most important practical outcome of book bans.

These findings prompt an expansion of the dominant censorship narrative around book bans. We do not propose that

conservative organizers are uninterested in restricting access to content they deem objectionable. However, our results demonstrate that bans are impractical efforts of censorship, insofar as they are directed at rather marginal cultural objects. Furthermore, at a time when roughly 97% of 3- to 18-year-olds have home internet access (38), it is unclear whether the removal of school books meaningfully restricts student access to their, or similar, content.

This raises the question: if they are not meaningful censorship campaigns, what are book bans accomplishing? We argue that our findings are suggestive evidence that book banning primarily serves as a reaction to increasingly contested, local political contexts. Given the strong association between conservatism and book bans in contemporary media coverage, it is somewhat surprising that the counties banning books are less conservative (as proxied by presidential elections) than neighboring counties, in particular since the 2016 election. One way to resolve this apparent contradiction is to interpret book bans as a form of collective action whose primary motive is to galvanize an apparently shrinking voting bloc by appealing to "culture war" antagonisms around race, gender, and sexual identity, rather than (or in addition to) as a form of control directed at access to certain cultural and intellectual goods. From this perspective, we identify censorship as a strategy potentially used to mobilize conservative voters, rather than an authoritarian, top-down approach of suppressing information in the perceived interest of the state.

In light of our findings, further work should better distinguish the political efficacy and spread of book bans, especially those targeting diverse casts of characters, women and nonbinary authors, and authors of color. In particular, identifying how book bans might galvanize conservatives' involvement in local politics and increase voter turnout will be required for better understanding the political effects of book bans. Our results are compatible with at least two different, but nonmutually exclusive interpretations that future work could disambiguate: (i) that due to the politically contested nature of these districts grassroots interest in local book bans precedes and ultimately benefits from the intervention of politicians and groups such as Moms for Liberty, or whether (ii) these organized groups identify candidate school districts on the basis of electoral politics and subsequently mobilize conservative parents in the area.

Additionally, while we find no evidence for mass censorship at a national scale, it is possible that book bans are associated with other outcomes at the local level. To this end, more qualitative work about the experience of parents and children in schools that ban books is necessary. Children's books, particularly those that win awards, already over-represent white characters (39) and there is a risk that further removing books featuring characters of color and LGBTQ+ characters from school shelves will only exacerbate what *Ebony* Elizabeth Thomas calls the "Diversity Crisis" in children's and young adult literature, whereby characters of color are scarce and often only depicted as the subjects of violent plot points (40). This could have serious implications for a child's sense of belonging in their community—regardless of whether they can still feasibly access the content of the books in other ways—that is invisible at the national level of our analysis. Even children belonging to social groups that are not targeted by these efforts may experience adverse consequences in learning outcomes if their schools become the sites of political contestation (41).

Our study is necessarily limited by data availability (Section S3). The PEN America Index of Book Bans is the most comprehensive resource available but should not be interpreted as an

absolute record of all bans. We are not able to differentiate between bans that are still in effect and those that were implemented and then overturned by the school board. The availability of books at each school is also not known: books might be placed on a no-purchase list or otherwise barred from acquisition before they have the opportunity to be banned from shelves; alternatively, banned books may never appear in the most conservative districts due to a lack of demand rather than a coordinated removal. In general, there is very little accessible data about book sales and interest. Despite the celebrity of “Best Seller” lists, book sales are heavily embargoed and it is not possible to extract usable sales data from Amazon, which represents about half of the on-line book sales and 75% of the ebook market (42). These data restrictions pose difficulties for assessing the state of banned books in particular and the diversity of the publishing industry in general (43). The open data portal provided by the Seattle Public Library offers a sign of promise for the collection of book interest and engagement data, although it requires a level of infrastructure unavailable to most school districts and libraries. It is possible that more robust and localized sales or library check-out data would be better positioned to identify the presence of a censorship effect, although our results suggest this is most likely not the case.

Our results allow us to better understand the rise of book bans. Book banning appears more similar to political strategies to receive attention and exert power. This is not to say that we should dismiss them as censorship attempts, but rather understand their primary purpose as most likely something other than information suppression, especially since the vast majority of the books banned are not popular books. In fact, the most sensational cases of book bans, which receive the majority of media attention, are rarely representative of the average banned book. While classic novels like Harper Lee’s *To Kill A Mockingbird* (44) and Toni Morrison’s *Beloved* (45) do appear in our sample of banned books, it is far more likely that banned books are picture books or contemporary educational nonfiction books about important historical figures. Attention should be directed towards the children’s books that make up the majority of the bans and future research should investigate which books are the target of bans and which stay on shelves.

As bans continue to increase across the country, our results suggest that these are political actions in addition to censorship tactics. The political ramifications of book bans remain under-examined. For example, in one Texas school district, an estimated \$30,000 was spent compensating hundreds of hours of staff time reviewing and adjudicating book bans during the 2022–2023 school year (46). As book bans continue, they will infringe upon student’s rights to information and incur heavy costs on taxpayers. Understanding their political context is an imperative.

Materials and methods

We rely on PEN America’s Index of School Book Bans (26), which includes instances in the United States in which student access to a book is restricted for a period of time, either in a school library or classroom. It is assembled through reviews of news articles, school websites, and letters to school districts, and should be considered a conservative estimate of book bans in the United States. It does not include books that were deaccessioned through standard procedures nor can it speak to books that were not purchased by the school in the first place. We use this dataset to identify each instance of book banning and each banned book (which could be banned by multiple school districts). The dataset documents a

total of 2,532 bans and 1,649 unique books in the United States during the 2021–2022 school year. Table S1 summarizes our data sources and their relation to our research questions.

County-level data

We matched each school district in the PEN America list to their respective counties and augmented each ban with county-level demographic data from the US Census Bureau. We combine this with the county-level presidential vote share data from the MIT Election Data and Science Lab (47) and data from the US Religion Census Religious Congregations and Membership Study (48). It is important to highlight that 29 school districts (out of 146) span more than one county. In these cases, all the counties that overlap with the school district were marked as a county that banned a book in the period. In our final sample, there are a total of 621 counties. Among these, 146 counties are home to school districts that enacted book bans during the specified period. The remaining 475 counties are counties in the same commuting zones as those that banned books but did not have their school districts enact book bans during that time.

Book-level metadata

We collect book-level metadata from multiple sources. First, we collected all the Goodreads genres listed for every banned book. Goodreads is a digital platform owned by Amazon that allows users to track their reading habits and leave reviews for books. Goodreads crowd-sources its genre labels through “user shelves” which function as a reader-produced classification system (49). For the 1,371 books in our sample that could be found on Goodreads with genre annotations (83%), there were a total of 143 unique genres, with each book having a maximum of 10 genre associations—such as “Law,” “Feminism,” “Young Adult,” or “Fantasy”—per book and an average of 7.2 genres per book before preprocessing. Because genre associations are derived from Goodreads users rather than publishers or authors, we manually created a set of genre correspondences to ensure qualitative consistency among genres (such that, e.g. a book tagged “Lesbian” would necessarily also be tagged “LGBT” if it were not already). After this preprocessing, the average book was linked to 7.8 genres.

With each book represented as a vector of genre dummy variables in this 143-genre feature space, we used the UMAP algorithm (50) to convert this sparse representation to a dense, 2D, and continuous one, then clustered these 2D representations of each book using the HDBSCAN algorithm (30). We combine these algorithms as UMAP’s dimensionality reduction has been shown to improve the performance of HDBSCAN (51, 52), while also enabling 2D visualization. As with many clustering applications, a model which yields too few clusters may obscure important variation in the data, while too many clusters can undermine the ultimate goal of summarizing data in a qualitatively legible way. Given that our purpose for clustering is to summarize and yield qualitative insights about our data rather than other downstream applications, we explored a range of hyperparameters and evaluated them in terms of the percent of books left unclustered (which we sought to minimize) and qualitatively, in terms of the perceived quality and distinctiveness of clusters. In the end we selected a model that yields six high-level genre clusters of books.

Finally, we used the Google Books API to gather short descriptions of each book (generally similar or identical to what appears on the book’s back cover). After removing blurbs from critics and author bios such that only descriptions of book content per se

remained, we fit a structural topic model (53) to these documents to provide an overview of lexical themes and their interrelations within the corpus of book descriptions. As the results of the topic model substantively replicate those of the genre-based analysis, we include them in the supplement (Fig. S1) rather than present them here.

Author demographic data

We collected the self-identified gender, race, and sexuality of each author in our dataset through an Amazon Mechanical Turk crowd-sourcing task that asked participants to collect such self-reported data from publicly available biographies on author websites, Wikipedia pages, and similar sources. We tested and timed the task to take around 3 min for a user new to the task to complete. We intentionally did not use a name-based algorithmic classifier to obtain this information because of known biases, especially for those who identify as people of color (54).

To assess the quality of information obtained, we audited a random sample of the results ($N = 50$ authors). We found that the majority of authors who self-identified their gender, race/ethnicity, or sexuality online were found by participants, but 22% of authors who self-identified as queer and 19% of authors who identified their race and/or ethnicity were not found by participants, so our estimates of the proportions of queer authors and authors of color are likely conservative. However, participants generally copied over the information accurately (98% accuracy for gender information, 85% for sexuality information and 100% for race/ethnicity information). Detailed results of this audit can be found in Table S3.

We compared our results to two reference datasets: (i) the proportion of writers and authors in the United States who listed their occupation as a writer or author in 2020, provided by the Bureau of Labor Statistics (31) and (ii) the proportion of authors who wrote the most popular books in the United States from 1950 to 2018 (32). Unfortunately, neither reference source collected cross-tabulated (intersectional) gender and race data (for example, the proportion of authors who are women of color).

Interest data

We leverage two distinct measures of interest to assess the possible impact of book bans, comparing the average interest in the 3 months prior to and following each ban. Outside of best seller lists, book sales data are heavily embargoed, with the top provider of book data to publishing houses refusing to license their data for academic research or interested individuals (43). To overcome this limitation, we negotiated access to restricted-use sales data from Bookshop.org, an online platform responsible for about 1% of the online book market in the United States. While sales data are likely the most robust measure of interest in a book, we complemented the sales data with a weaker yet broader measure of interest—search data from Google Trends (for more details see Section S3). The Bookshop.org data is normalized as they shared the data with us under the condition that we do not report exact sales. Both the Bookshop.org and Google Trends results reported in the main text are measures of interest at the national level. We also ran Google Trends results at the state level and found that interest did not significantly change post-ban (Fig. S9). Finally, the most granular level of interest is at the local level, and there is very little public data available at the local level. We ran the same analysis using Seattle Library's open data portal, where we can collect the number of checkouts for each book in the city's library system, and again found no significant change

(Fig. S10). However, Seattle is not representative of the regions that typically ban books, thus we cannot draw specific local-level conclusions, but as a large metropolitan area we expect that national-level effects, were they to exist, should be visible in these data and thus it serves as an additional robustness check of our national-level results.

For each interest indicator, we conduct a pre-post design, comparing the average interest across the 3 months preceding the ban with the three following months. The latter group contains the ban month itself. We chose a monthly time series measure because the PEN America dataset includes the month of each ban but not the day. Indeed, bans are likely to occur over several weeks as a group petitions the school district for the removal of a book, meetings are held, and a final decision is made. For robustness, we ran the same tests on different time groups (1 month, 4 months, 6 months); all groupings produce similar, nonsignificant results (Fig. S7).

Acknowledgments

We thank Chris Bail, Lizzie Martin, Jay A. Pearson, Alejandra Regla-Vargas, Nina Wang, Sam Zhang, and reviewers for helpful comments. We additionally thank Bookshop.org for their cooperation and generous sharing of data. This manuscript was posted on a preprint: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4618699.

Supplementary Material

Supplementary material is available at PNAS Nexus online.

Funding

This research was assisted by a Social Science Research Council (SSRC)/Summer Institutes in Computational Social Science Research Grant.

Author Contributions

All authors contributed equally to conceptualization, methodology, formal analysis, data curation, and writing.

Data Availability

Open-source code used for our analyses is available at <https://zenodo.org/records/10982953>. All underlying source data used to run our analyses is available at <https://zenodo.org/records/10982955>, with the exception of the restricted-use book sales data and the author demographic data. Anonymized versions of the book sales data and author demographic data are included in the open-source repository, but the full versions may be available upon request to qualified researchers.

References

- 1 Zimmerman J. 2022. *Whose America? Culture wars in the public schools*. 2nd ed. Chicago (IL): The University of Chicago Press.
- 2 Boyer PS. 2009. Gilded-age consensus, repressive campaigns, and gradual liberalization. In: Kaestle C, Radway J, editors. *A history of the book in America: volume 4: print in motion: the expansion of publishing and reading in the United States, 1880–1940*. Chapel Hill (NC): University of North Carolina Press. p. 276–298.

- 3 Donaldson S. 1991. Censorship and a farewell to arms. *SAF*. 19(1): 85–93.
- 4 Oltmann SM. 2023. *The fight against book bans: perspectives from the field*. London, England: Bloomsbury.
- 5 Caldweall-Stone D. 2022. Letter to house oversight committee opposing book bans and challenges to free speech. <https://alair.ala.org/handle/11213/18004>
- 6 Curtis J. 2022. The effect of the 2020 racial justice protests on attitudes and preferences in rural and urban America. *Soc Sci Q*. 103(1):90–107.
- 7 Yuracko KA. 2022. The culture war over girls' sports: understanding the argument for transgender girls' inclusion. *Villanova Law Rev*. 67(4):717–758.
- 8 Gabbatt A. 2022. 'Unparalleled in intensity'—1,500 book bans in US school districts. *The Guardian*. <https://www.theguardian.com/us-news/2022/apr/07/book-bans-pen-america-school-districts>
- 9 Harris EA, Alter A. 2022. With rising book bans, librarians have come under attack. *The New York Times*. <https://www.nytimes.com/2022/07/06/books/book-ban-librarians.html>
- 10 Harris EA, Alter A. 2023. Book removals may have violated students' rights, education department says. *The New York Times*. <https://www.nytimes.com/2023/05/22/books/book-banning-education-civil-rights.html>
- 11 Natanson H, Rozsa L. 2022. Students lose access to books amid 'state-sponsored purging of ideas.' *Washington Post*. <https://www.washingtonpost.com/education/2022/08/17/book-ban-restriction-access-lgbtq/>
- 12 Delgado R, Stefancic J, Harris A. 2017. *Critical race theory*. 3rd ed. New York (NY): NYU Press.
- 13 Citizens for Renewing America. 2021. Combatting critical race theory in your community: an A to Z guide on how to stop critical race theory and reclaim your local school board. <https://citizensrenewingamerica.com/issues/combating-critical-race-theory-in-your-community/>
- 14 Lavietes M. 2023, April 25. Over half of 2022's most challenged books have LGBTQ themes. *NBC News*. <https://www.nbcnews.com/nbc-out/out-politics-and-policy/half-2022s-challenged-books-lgbtq-themes-rcna81324>
- 15 Harris EA, Alter A. 2022. A fast-growing network of conservative groups is fueling a surge in book bans. *The New York Times*. <https://www.nytimes.com/2022/12/12/books/book-bans-libraries.html>
- 16 Vissing Y, Juchniewicz M. 2023. Children's book banning, censorship and human rights. In: Zajda J, Hallam P, Whitehouse J, editors. *Globalisation, values education and teaching democracy*. Cham (Switzerland): Springer International Publishing. p. 181–201.
- 17 Perry A. 2023. Pico, LGBTQ+ book bans, and the battle for students' first amendment rights. *Tul. JL & Sexuality*. 32:197–219.
- 18 Shearer M. 2022. Banning books or banning BIPOC? *Nw U L Rev Online*. 117:24–45.
- 19 Jaeger PT, et al. 2022. Exuberantly exhuming McCarthy: confronting the widespread attacks on intellectual freedom in the United States. *Libr Q*. 92(4):321–328.
- 20 Ferguson CJ. 2014. Is reading "banned" books associated with behavior problems in young readers? The influence of controversial young adult books on the psychological well-being of adolescents. *Psychol Aesthet Creat Arts*. 8(3):354–362.
- 21 Hobbs WR, Roberts ME. 2018. How sudden censorship can increase access to information. *Am Political Sci Rev*. 112(3):621–636.
- 22 Roberts ME. 2018. *Censored: distraction and diversion inside China's Great firewall*. Princeton (NJ): Princeton University Press.
- 23 Roberts ME. 2020. Resilience to online censorship. *Ann Rev Political Sci*. 23(1):401–419.
- 24 Nabi Z. 2014. ~~Resistance~~ censorship is futile. *First Monday*. <https://doi.org/10.5210/fm.v19i11.5525>
- 25 Swenson A. 2023. Moms for Liberty rises as power player in GOP politics after attacking schools over gender, race. *Associated Press*. <https://apnews.com/article/moms-for-liberty-2024-election-republican-candidates-f46500e0e17761a7e6a3c02b61a3d229>
- 26 PEN America. 2022. Pen America index of school book bans- 2021–2022 [dataset]. <https://pen.org/banned-book-list-2021-2022/>
- 27 Morozov E. 2011. *The net delusion: the dark side of internet freedom*. New York (NY): Public Affairs.
- 28 Jansen SC, Martin B. 2015. The streisand effect and censorship backfire. *Int J Commun*. 9(0):Article 0.
- 29 McInnes L, Healy J, Melville J. 2020. UMAP: uniform manifold approximation and projection for dimension reduction. arXiv:1802.03426. <https://doi.org/10.48550/arXiv.1802.03426>, preprint: not peer reviewed.
- 30 McInnes L, Healy J, Astels S. 2017. Hdbscan: hierarchical density based clustering. *J Open Source Softw*. 2(11):205.
- 31 Bureau of Labor Statistics. 2022. *Employed persons by detailed occupation, sex, race, and Hispanic or Latino ethnicity: U.S. Bureau of Labor Statistics*. <https://www.bls.gov/cps/cpsaat11.htm>
- 32 So RJ. 2020. *Redlining culture: a data history of racial inequality and postwar fiction*. New York (NY): Columbia University Press.
- 33 Blalock HM. 1967. *Toward a theory of minority-group relations*. New York: Wiley.
- 34 Blumer H. 1958. Race prejudice as a sense of group position. *Pac Sociol Rev*. 1:3–7.
- 35 Nelson AA, Gazley B. 2014. The rise of school-supporting non-profits. *Educ Finance Policy*. 9(4):541–566.
- 36 Droubay BA, Butters RP, Shafer K. 2021. The pornography debate: religiosity and support for censorship. *J Relig Health*. 60(3):1652–1667.
- 37 Kobabe M. 2019. *Gender queer: a memoir*. St. Louis (MO): Lion Forge.
- 38 National Center for Education Statistics. 2023. Children's Internet Access at Home. US Department of Education, Institute of Education Sciences. <https://nces.ed.gov/programs/coe/indicator/cch/home-internet-access#suggested-citation>
- 39 Adukia A, Eble A, Harrison E, Runesha HB, Szasz T. 2021. What we teach about race and gender: representation in images and text of children's books (Working Paper 29123). *National Bureau of Economic Research*.
- 40 Thomas EE. 2019. *The dark fantastic: race and the imagination from Harry Potter to the hunger games*. New York (NY): New York University Press.
- 41 Burmester S, Howard LC. 2022. Confronting book banning and assumed curricular neutrality: a critical inquiry framework. *Theory Pract*. 61(4):373–383.
- 42 Anderson P. 2020, August 17. US publishers, authors, booksellers call out Amazon's "Concentrated Power." *Publishing perspectives*. <https://publishingperspectives.com/2020/08/us-publishers-authors-booksellers-call-out-amazons-concentrated-power-in-the-book-market/>
- 43 Walsh M. 2022, October 4. Where is all the book data? *Public books*. <https://www.publicbooks.org/where-is-all-the-book-data/>
- 44 Lee H. 2002. *To kill a mockingbird*. 1st Perennial Classics ed. Philadelphia (PA): HarperCollins.
- 45 Morrison T. 1987. *Beloved*. New York (NY): Alfred A. Knopf, Inc. Beloved.
- 46 Ryan S. 2023, March 29. More than \$30K of taxpayers' money, 220 hours spent on single Spring Branch ISD book ban, docs show. *ABC13 Houston*. <https://abc13.com/spring-branch-isd-book-ban-school-library-books-student-resources-texas-schoolbook-restrictions/13037457/>

- 47 MIT Election Data and Science Lab. 2022. *County presidential election returns 2000–2020* (Version 11) [dataset]. *Harvard Dataverse*. <https://doi.org/10.7910/DVN/VOQCHQ>
- 48 Jones DE, Houseal R, Krindatch A, Stanley R, Grammich C, Hadaway K, Taylor RH. 2018. *U.S. religion census—religious congregations and membership study, 2010* (State File). <https://doi.org/10.17605/OSF.IO/X8D69>
- 49 Walsh M, Antoniak M. 2021. The goodreads “classics”: a computational study of readers, Amazon, and crowdsourced amateur criticism. *J Cult Analyt*. 6(2):243–287. <https://doi.org/10.22148/001c.22221>.
- 50 McInnes L, Healy J, Saul N, Großberger L. 2018. UMAP: uniform manifold approximation and projection. *J Open Source Softw*. 3(29):861.
- 51 Allaoui M, Kherfi ML, Cheriet A. 2020. Considerably improving clustering algorithms using UMAP dimensionality reduction technique: a comparative study. In: El Moataz A, Mammass D, Mansouri A, Nouboud F, editors. *Image and signal processing*. Cham (Switzerland): Springer International Publishing. p. 317–325.
- 52 Blanco-Portals J, Peiró F, Estradé S. 2022. Strategies for EELS data analysis. Introducing UMAP and HDBSCAN for dimensionality reduction and clustering. *Microsc Microanal*. 28(1):109–122.
- 53 Roberts ME, Stewart BM, Tingley D. 2019. Stm: an R package for structural topic models. *J Stat Softw*. 91:1–40.
- 54 Lockhart JW, King MM, Munsch C. 2023. Name-based demographic inference and the unequal distribution of misrecognition. *Nat Hum Behav* 7(7):1084–1095.